

Slide 1 of 10

Title Slide: SASS Sample Design, Weights, Variance, and Missing Data

Slide 2 of 10

SASS uses a complex design for selecting the sample from which data were collected. As discussed in the common modules, when analyzing data from sample surveys, certain procedures must be used to assure that estimates made from the data are representative of the population and that hypothesis tests are accurate. Specifically, the final weight must be applied and standard errors must be calculated using appropriate procedures. This module discusses these issues specifically in relation to analyses of data from SASS. These same procedures can also be used for the Teacher Follow-up Survey (TFS) and Principal Follow-up Survey (PFS) data, which will be discussed in detail in the module titled “Follow-up Surveys to SASS”.

Slide 3 of 10

The sampling frame for SASS is based upon two universe collections of elementary and secondary schools in the United States, the Common Core of Data (CCD) for traditional public and public charter schools, and the Private School Universe Survey (PSS) for private schools. CCD is based on survey data collected annually by the National Center for Education Statistics (NCES) from each state education agency. PSS is based upon lists of private school membership organizations and is updated through a survey sent out by NCES every other year.

The availability of information about schools’ characteristics from the sampling frame enables a sample design to be created ahead of data collection. The SASS two-stage sampling design involves the selection of a nationally representative sample of elementary and secondary education schools, followed by the sampling of teachers within these schools. SASS is explicitly stratified by school sector – that is, sampled schools explicitly represent all major sectors of elementary and secondary education, including traditional public, private, and public charter schools.

Within these school sectors the sample is stratified by state for public schools, by affiliation type for private schools, and at the national level for public charter schools. Public charter schools are not distributed equally around the United States. A relatively large proportion of the nation’s public charter schools are clustered in a few states. Selecting a sample of public charter schools within the larger population of public schools is achieved in SASS by taking a higher proportion of public charter schools out of that overall school population. SASS sample design supports estimates for traditional public schools at the state level, for public charter schools at the national level, and for private schools at the affiliation type level. Comparisons of traditional public and public charter schools or of public and private schools can only be supported at the national level. Caution should be used in generalizing beyond these sampling strata.

Slide 4 of 10

The second stage of sampling in SASS is selecting a sample of teachers within the sampled school. From the sampled schools, a roster of all teachers in the school is requested. The list of teachers also includes information on the subject matter taught, the grade level taught, whether the teacher is full-time or part-time, and years of teaching experience. Because NCES does not have a universe survey of public school or private school teachers, this second stage of sampling for SASS is an operation that takes place when the school-level questionnaires are sent to the sampled schools.

Stratifying schools or teachers makes sampling more efficient. When there are many small sized schools, only a few have to be selected to be representative of all smaller sized schools. Achieving a geographically-balanced sample within a stratum is possible by sorting all of the schools on one or more characteristics, such as community size, within the direct sampling strata.

Once teachers were sorted into various groups within the school, the probabilities of selecting any teacher within a given group were set at an equal level. The clustering of teachers within specific grade level ranges of schools helps achieve the sampling goals of representing all types of teachers – new and experienced, with diverse teaching assignments, part-time as well as full-time – while minimizing how many teachers are selected at any particular school.

Slide 5 of 10

As was discussed in the Common Module “Analyzing NCES Complex Survey Data,” weights should be used in analyses of data from surveys such as SASS in order to make estimates produced from the sample representative of the target population. In analyses using SASS data, the sampling weights are designed to account for unequal probabilities for selection. The final weights in SASS assigned to each respondent incorporate all stages of sampling, as applicable, and all types of sampling and nonresponse adjustment. Sampling weights must be used in order to bring the sample estimate into the correct proportion for the target population.

Each data file in SASS has a specific final weight that reflects all stages of sampling (if more than one is applicable), plus sampling adjustments that occur after the sample was selected, and also for unit-level nonresponse.

Each SASS component final weight (school, district, principal, teacher, and school library media center) has its own alphabetic prefix. The final weight for a specific component level uses this prefix, plus the letters FNLWGT. Thus, the final weights in SASS are: SFNLWGT (school); DFNLWGT (district); AFNLWGT (principal or administrator); TFNLWGT (teacher); and MFNLWGT (library media center).

For information about weighting, please view the Common Module titled “Analyzing NCES Complex Survey Data,” if you have not already done so. This module can be accessed by clicking on the underlined screen text ‘weights.’

Slide 6 of 10

For every respondent in SASS, the final weight is provided for analysis using data from all applicable stages of sampling and nonresponse adjustment. SASS weights on a given data file sum across the sample cases to the population of that component type (districts, principals, schools, teachers, or library media centers) in 2011-12. Every SASS weight is adjusted for nonresponse. Estimates produced using these nonresponse-adjusted weights are representative of the characteristics of the population of the particular component in SASS – districts, principals, schools, teachers, or library media centers even though some of the sampled potential respondents did not participate in the survey.

Slide 7 of 10

Development of the SASS final weights reflect the cross-sectional design of the SASS survey system. That is, every data collection of SASS is sampled independently, even for schools that may fall into sample for more than one data collection. Every sampled potential respondent for a given SASS data collection has only one probability of selection. Therefore, there is a single final weight that contains all of the sampling and nonresponse adjustment factors applicable to the respondent.

Weights matter in SASS because in computing estimates without the use of the final weight, any statistical software package is assuming a *simple random sample* (SRS) with an equal probability of selection for each respondent. SASS does not use an equal probability of selection, so *the final weight value varies from one respondent record to another*. For example, new teachers are oversampled, because new teachers are a relatively small proportion of the teacher workforce. If a simple random sample were to be used in SASS, then there might not be a large enough selection of new teachers to support estimates at the state or affiliation type level. Generating an estimate from SASS data without the use of the final weight yields a result that is not accurate because it does not take into account the unequal probabilities of selection used in SASS. In unweighted estimates of teachers, new teacher data would be over-represented relative to the overall population of teachers. Applying the teacher final weight allows new teacher data to be put into correct proportion among all teachers.

Slide 8 of 10

In the common module titled, “Statistical Analysis of NCES Datasets Employing a Complex Sample Design,” two standard error calculation procedures were discussed: Replication Techniques and Taylor Series linearization.

Replication is a method that calculates appropriate standard errors based on differences between estimates from the full sample and a series of created subsamples, or replicates. In this method, you need to select replicate weights that are associated with the particular SASS data file’s final weight. Every final weight in SASS has a set of replicate weights that are associated with it. For example, the replicate weights associated with the final sampling weight TFNLWGT are TREPWT1 (spelled T-R-E-P-

SASS Sample Design, Weights, Variance, and Missing Data

W-T-1) through TREPWT88 (spelled T-R-E-P-W-T-88). It is important to include the entire set of replicate weights in the weight statement.

SASS replication weights use a Balanced Repeated Replication (or BRR) method so that is the method that should be specified within the statistical software used for analysis, particularly the bootstrap option.

It is important to note that Balanced Repeated Replication is the only accurate method for calculating standard errors in SASS, as no Primary Sampling Unit or PSU codes and strata identifiers, which are required for Taylor series linearization, are provided in the SASS data file.

For information about standard error calculation more generally, please view the Common Module titled “NCES Datasets Employing a Complex Sample Design,” if you have not already done so. This module can be accessed by clicking on the underlined screen text ‘calculates appropriate standard errors.’

Slide 9 of 10

In addition to the resource documents provided in the Common Module titled, ‘Statistical Analysis of NCES Datasets Employing a Complex Sample Design,’ this Resource Document provides examples of SUDAAN code to calculate standard errors for different types of SASS data. Continuous data estimates, such as the total student enrollment, require PROC DESCRIPT. Percentage estimates require PROC RATIO. The resource document is accessible by clicking on the underlined screen text, ‘SUDAAN code.’ It is important to note that using this code for your own research purposes will necessitate substituting your variables of interest.

Slide 10 of 10

This module provided information about the sampling design, weights and variables associated with calculating estimates and standard errors for SASS data. This module detailed how weights must be applied and standard errors must be calculated to ensure that estimates made from the data are representative of the population and that hypothesis tests are accurate specifically in relation to analyses of data from SASS.

Important resources that have been provided throughout the module are summarized in this slide along with the module’s objectives for your reference.

You may now proceed to the next module in the series, or click the exit button to return to the landing page.